DOCUMENT RESUME

ED 18[illegible] 765 EC 124 305

AUTHOR Greenleaf, Wayne: Smith, Henry L.
TITLE Non-Discriminatory Intellectual Assessment Procedures.
INSTITUTION Louisiana State Dept. of Education, Baton Rouge.
PUB DATE 78
NOTE 17p.: Print in tables is marginally legible. Prepared through the Division of Special Educational Services.

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Culture Fair Tests: *Disadvantaged: Elementary Secondary Education: *Intelligence Tests: Minority Groups: *Racial Differences: *Test Bias: *Testing: Test Validity
IDENTIFIERS *System of Multicultural Pluralistic Assessment: Wechsler Intelligence Scale for Children (Revised)

ABSTRACT

The paper discusses test bias: presents the rationale of the System of Multicultural Pluralistic Assessment (SOMPA), a test instrument which assumes equal cultural and racial intellectual potential but asserts that there is not one normal curve of measured intellectual performance: and suggests that the test would correct racial and cultural test bias in Louisiana. Evidence of test bias has been demonstrated in the Wechsler Intelligence Scale for Children-Revised based on traditional test bias definitions. Reports of preliminary studies have shown that the SOMPA test has brought racial and cultural balance to systems which previously had overrepresented minorities. Steps in securing an intellectual assessment of estimated learning potential are outlined. (PHR)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED189765

U S DEPARTMENT OF HEALTH,
EDUCATION & WELFARE
NATIONAL INSTITUTE OF
EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

STATE OF LOUISIANA
DEPARTMENT OF EDUCATIO

NON-DISCRIMINATORY INTELLECTUAL ASSESSMENT PROCEDURES

DIVISION OF SPECIAL EDUCATIONAL SERVICES

BUREAU OF PUPIL APPRAISAL

This public document was published at a cost of $.33 per copy by the Printing Section, Department of Education, P. O. Box 44064, Baton Rouge, Louisiana 70804, under authority of special exception by the Division of Administration. This material was printed in accordance with the standards for printing by state agencies established pursuant to R.S. 43:31.

J. Kelly Nix
State Superintendent of Education

1978

EC124305

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

Wayne Greenlaw

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

ACKNOWLEDGMENT

Prepared by Dr. Wayne Greenleaf, Director of the Bureau of Pupil Appraisal and Dr. Henry L. Smith, Assistant Superintendent of Special Educational Services.

Appreciation is expressed to Dr. Tom Staats, Director of Pupil Appraisal Center, Caddo Parish, who prepared the first draft.

Thanks to Dr. Jane Mercer for permission to reproduce portions of her SOMPA Workbook and for long discussions about her unpublished research.

Credit must also be given to staff members of the Division who read the rough draft and made valuable suggestions.

Act 754 of the 1977 Louisiana Legislature mandates a multidisciplinary, individual evaluation for each child prior to placement in special educational programs. The Regulations adopted by the Board of Elementary and Secondary Education for Act 754 specify the clear intent of the State of Louisiana to provide the most accurate and fair assessment possible to each individual 0-21 years of age who is suspected of being exceptional. Two statements from the Regulations specify a portion of the standards for implementing this intent as it relates to various forms of possible discrimination in evaluation.

§436(b) Tests and other evaluation materials used to assess the nature and extent of a suspected exceptionality and to assess general or specific areas of educational need, must be:

. . .

(4) Free of racial, cultural, language or sex bias;

§436(c) No individual or group IQ tests may be administered as part of any individual evaluation, screening or assessment unless the test has been approved for the proposed use by the State Board after a review of a formal request by the State Superintendent for such approval. The formal request must be accompanied by a report signed by the Superintendent which sets forth in detail and convincingly demonstrates that: (1) the tests,

with respect to the use or uses proposed (for the racial and cultural groups specified) are free of racial and cultural bias, and (2) have been standardized with and are reliable and valid (for the uses specified) with respect to children of the specified racial and cultural groups.

In reality there are no known educational or psychological tests which have been demonstrated to be totally free of cultural and racial bias or discrimination.

However, a new psychometric and sociometric instrument called the System of Multicultural Pluralistic Assessment (SOMPA) authored by Dr. Jane Mercer and June Lewis has been published recently by the Psychological Corporation. In standardization studies and field trials SOMPA appears to have corrected the racial, cultural and socioeconomic bias inherent in the Wechsler Intelligence Scale for Children - Revised (WISC-R).

This paper will be organized to discuss definitions of test bias, rationale of SOMPA and then the recommendations for Louisiana. All Figures are from Jane Mercer's SOMPA training manual and are reproduced with her permission.

Definition of Test Bias

Webster's dictionary describes bias as: "partiality, a tendency to favor unfairly, inequitable."

In traditional psychometric usage, bias is usually determined by test items which show partiality to one group by including a majority of questions from the cultural heritage or experience of that group. Bias is also present if inferences about "intelligence" or "potential" are made on the basis of test scores which consistently demonstrate a higher or lower error in comparison of performance of different cultural groups (see Figure 1 on the following page).

Available data which relate to the six areas which evidence cultural bias will be summarized in the following portion of this section as it related to the Wechsler Intelligence Scale for Children. Much of this evidence of cultural bias was presented in the Larry P. test case in California.

Figure 2 shows that on the WISC-R there are three statistically significantly different means and distributions in scores, suggesting test item bias.

FIGURE 2

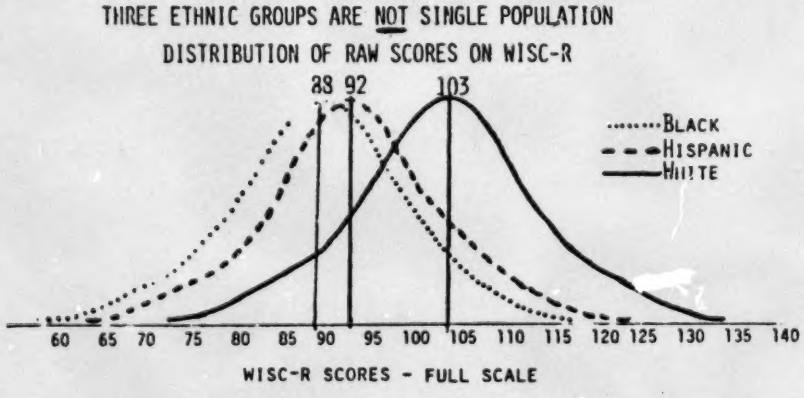

THREE ETHNIC GROUPS ARE NOT SINGLE POPULATION
DISTRIBUTION OF RAW SCORES ON WISC-R
88
92
103
BLACK
HISPANIC
WHITE
60 65 70 75 80 85 90 95 100 105 110 115 120 125 130 135 140
WISC-R SCORES - FULL SCALE

FIGURE 1

BIAS OR FAIRNESS IN TESTING

CUSTOMARY USAGE

WEBSTER'S DICTIONARY: BIAS = PARTIALITY, TENDENCY TO FAVOR UNFAIRLY, INEQUITABLE

A TEST IS BIASED IF IT:

1. SHOWS PARTIALITY TO ONE GROUP BY INCLUDING MAINLY QUESTIONS FROM THEIR CULTURAL HERITAGE AND FEW IF ANY QUESTIONS FROM THE CULTURAL HERITAGE OF OTHER GROUPS AND
2. INFERENCES ABOUT 'INTELLIGENCE' OR 'POTENTIAL ARE MADE ON BASIS OF TEST SCORES BY INEQUITABLE COMPARISON OF PERFORMANCE OF DIFFERENT CULTURAL GROUPS

EVIDENCE OF SUCH CULTURAL BIAS:

1. EXAMINATION OF TEST ITEMS SHOWS QUESTIONS, LANGUAGE, AND SKILLS FROM ONLY ONE CULTURE.
2. SCORES ON TEST ARE HIGHLY CORRELATED WITH CULTURAL BACKGROUND CHARACTERISTICS
3. DIFFERENCES IN AVERAGE SCORES ON TEST ARE SIGNIFICANT AND FAVOR THE GROUP WHOSE CULTURAL HERITAGE IS REPRESENTED IN TEST.
4. MINORITY GROUP MEMBERS MOST ASSIMILATED TO CULTURE OF THE TEST SCORE BETTER THAN THOSE LESS ASSIMILATED.
5. EXPERIMENTAL STUDIES DEMONSTRATE THAT MINORITY CHILDREN SOCIALIZED TO CULTURE OF THE TEST DO BETTER THAN THOSE WHO DO NOT EXPERIENCE SUCH SOCIALIZATION. (HEBER-GARBER)
6. MINORITY CHILDREN ADOPTED INTO HOMES OF PERSONS FROM CULTURAL BACKGROUND REPRESENTED IN THE TEST DO BETTER ON THE TEST THAN THOSE NOT ADOPTED. (SCARR-WEINBERG)

Figures 3 and 4 indicate that as more sociocultural factors are controlled (i.e., the more a group has been successful in American culture) the more the Hispanic and black I.Q. means approach the white dominant or middle class mean for the I.Q. test (Mercer, 1974).

FIGURE 3

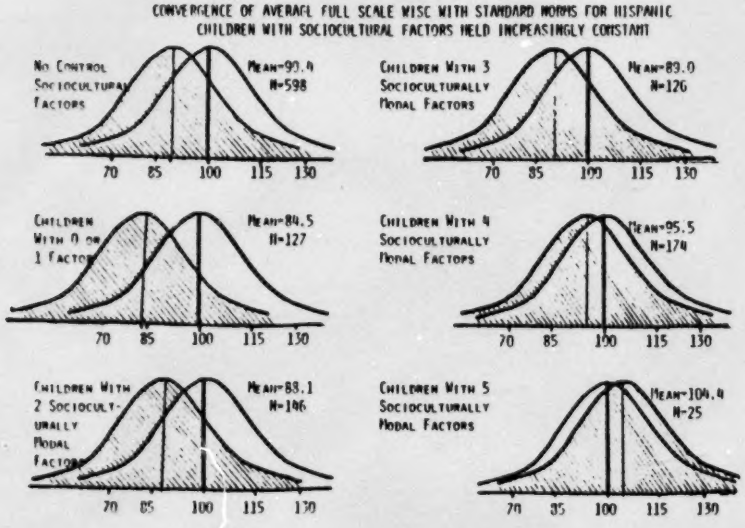


FIGURE 4

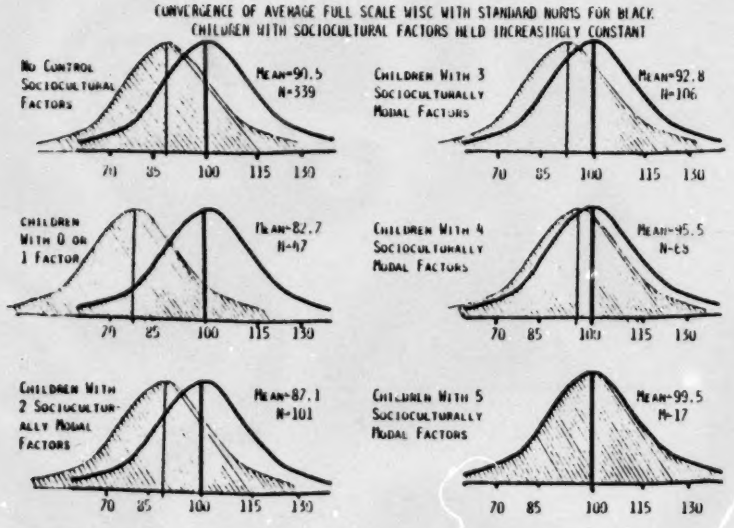


Figure 5 shows data from the Heber and Garber 1974 studies of the effects of infant development stimulation on the I.Q.'s of disadvantaged ghetto children in comparison to matched controls not receiving such stimulation and socialization. The Scarr-Weinberg (1975) study (Figure 6) shows that black children adopted into white core culture homes have statistically higher I.Q.'s and reading scores than matched controls raised in minority homes. Taken together these results strongly suggest that learning and sociocultural experience are part of what traditional I.Q. tests are measuring. In fact, Mercer (1978) has demonstrated that 27.5% of the WISC-R Verbal I.Q. variance, 13.7% of the Performance I.Q. variance and 24.3% of the Full Scale I.Q. variance can be accounted for by sociocultural factors (amount of exposure to core culture).

All of the above data demonstrate that bias is present according to the previous definitions when measuring intelligence using the Wechsler scales.

FIGURE 5

MILWAUKEE STUDY

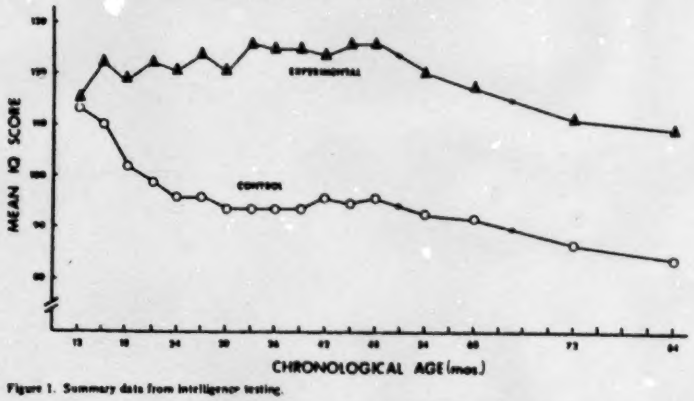
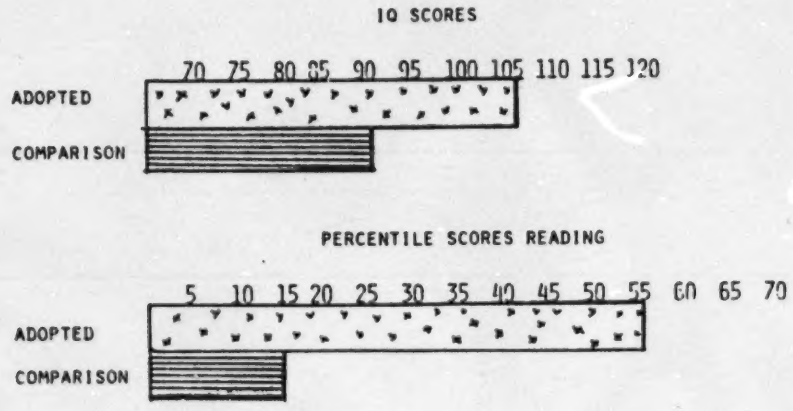


Figure 1. Summary data from intelligence testing.

GARBER, HOWARD L., INTERVENTION IN INFANCY, IN REGAB AND RICHARDSON (ED.), THE MENTALLY RETARDED AND SOCIETY: A SOCIAL SCIENCE PERSPECTIVE, UNIVERSITY PARK PRESS, BALTIMORE, 1974.

FIGURE 6

BLACK CHILDREN ADOPTED INTO CORE CULTURE HOMES


IQ SCORES
70 75 80 85 90 95 100 105 110 115 120
ADOPTED
COMPARISON
PERCENTILE SCORES READING
5 10 15 20 25 30 35 40 45 50 55 60 65 70
ADOPTED
COMPARISON

BASED ON FINDINGS REPORTED BY SANDRA SCARR AND RICHARD A. WEINBERG, WHEN BLACK CHILDREN GROW UP IN WHITE HOMES, PSYCHOLOGY TODAY, DECEMBER, 1975.

Psychometricians have a completely different way of looking at test bias as a statistical phenomenon. One of the most widely accepted statistical definitions of test bias is the Cleary Definition and its corollaries listed in Figure 7 on the following page.

Dr. Mercer (1978) chose to look at the prediction equations (regression equations predicting racially different student role performance (grade point average)) from WISC-R scores the previous year. Figures 8 and 9 show that both the slopes and points of origin of the regressions are quite different between racial groups and, therefore, bias is present according to the Cleary Definition, if grade point average is used as the criterion variable.

FIGURE 7

BIAS OR FAIRNESS IN TESTING

STATISTICAL - PSYCHOMETRIC USAGE

CLEARY'S DEFINITION: "A TEST IS CONSIDERED FAIR FOR A PARTICULAR USE IF THE INFERENCE DRAWN FROM THE TEST SCORE IS MADE WITH THE SMALLEST FEASIBLE RANDOM ERROR AND IF THERE IS NO CONSTANT ERROR IN THE INFERENCE AS A FUNCTION OF MEMBERSHIP IN A PARTICULAR GROUP." (AMERICAN PSYCHOLOGIST, JAN. 1975)

A TEST IS BIASED OR UNFAIR:

1. IF THE VARIABILITY AROUND THE REGRESSION LINE I.E. RANDOM ERROR, IS NOT SIMILAR FOR BOTH GROUPS.
2. IF THE LEVEL OF THE REGRESSION LINE IS NOT SIMILAR FOR BOTH GROUPS AND HENCE PRODUCES A CONSTANT ERROR IN PREDICTION.
3. IF THE SLOPES OF THE REGRESSION LINES ARE NOT SIMILAR AND HENCE PRODUCE DIFFERENT PREDICTIONS FOR EACH GROUP.

EVIDENCE OF BIAS OR UNFAIRNESS:

1. THE SIZE OF THE CORRELATION COEFFICIENTS IS NOT SIMILAR.
2. THE INTERCEPTS FOR THE REGRESSION LINES ARE NOT SIMILAR.
3. THE SLOPES OF THE REGRESSION LINES ARE NOT PARALLEL.

NOTE: UNDER THIS DEFINITION, A DIFFERENCE IN THE AVERAGE SCORES OF THE TWO GROUPS IS NOT ACCEPTED AS AN INDICATOR OF "UNFAIRNESS" OR "BIAS".

FIGURE 8

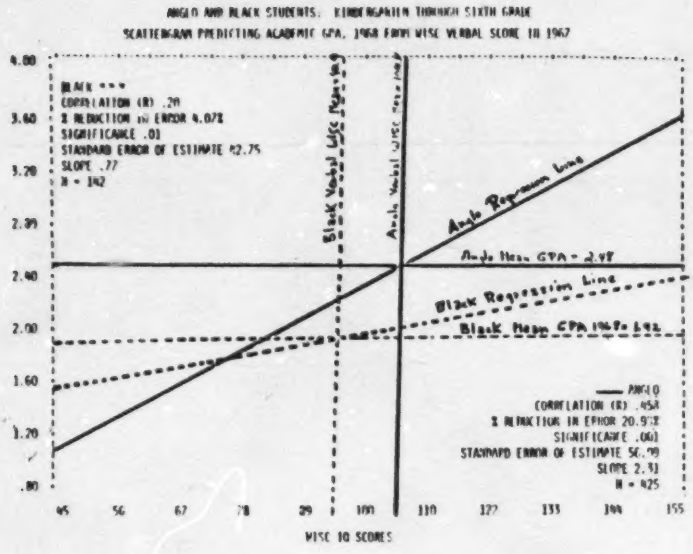
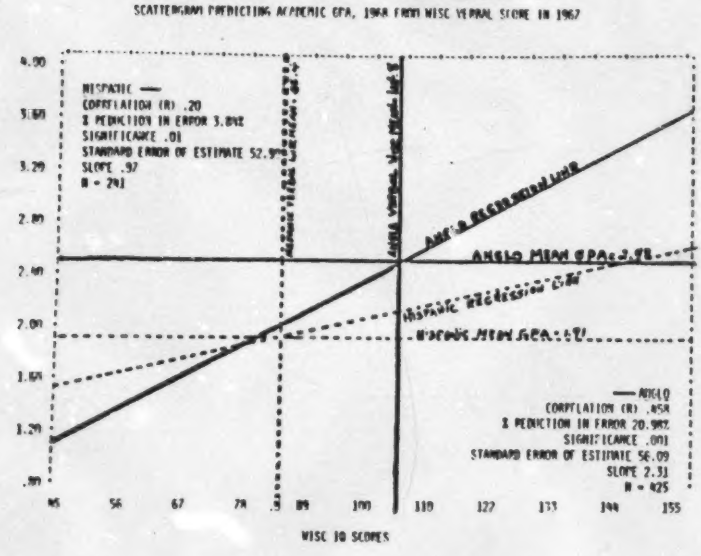


FIGURE 9



If achievement scores, not grades, are the criterion variable, the tests are not biased by this definition. This is the argument of the traditional I.Q. test proponents. No one denies that present I.Q. tests predict achievement fairly well, but Dr. Mercer proposes that achievement and I.Q. tests are really measuring the same thing (spuriously high correlation due to shared item variance). Other experts disagree. It is important to note, however, that Dr. Mercer's position is partially substantiated by longitudinal studies of life success which suggest that number of years completed successfully in regular education is a better predictor of life success than I.Q. (therefore, "G.P.A. and passing" are more relevant than achievement scores).

The position taken by the Louisiana State Department of Education is that both traditional I.Q. and SOMPA Corrected I.Q.'s (Estimated Learning Potentials) must be looked at, traditional I.Q.'s to identify academically "at risk" youngsters for regular education and SOMPA Estimated Learning Potential Scores to structure special education placements and expectancies.

SOMPA Rationale and Cultural Correction

The System of Multicultural Pluralistic Assessment assumes that all cultural and racial groups have equal intellectual potential but that there is not one normal curve of measured intellectual performance. When measured performance is used there is an infinite number of normal curves whose means differ as a function of the amount of core Anglo culture exposure and socialization. Therefore, the System uses a set of sociocultural

scales to place every youngster on the appropriate normal curve for youngsters he can fairly be compared with because they have equal core culture exposure (i.e., the sociocultural variance or influence is controlled). Thus, according to this system, an unbiased estimate of his learning potential can be made.

In addition to providing culturally corrected scores, the SOMPA offers an excellent Adaptive Behavior Measure for satisfying the two-part definition of mental retardation recommended by the American Association of Mental Deficiency and the State Department of Education. The best feature of the SOMPA Adaptive Behavior Measure is that it has a veracity scale which measures the validity of information given by the informant.

SOMPA has not been used enough to judge its full worth, but preliminary data look encouraging. Fisher (1978) in the Corpus Christi Independent School District found that SOMPA ELP and Adaptive Behavior Scale use brought racial balance in Educable Mentally Retarded classes in which Mexicans and blacks were previously over-represented. He also found no significant cultural differences in the means of the other SOMPA measures. A similar study in the Pueblo, Colorado, public school system resulted in essentially the same findings in 1977 (Personal Communication, Mercer).

Reschley (1978) found no differences in ELP means from the California samples when he tested Anglos, Blacks, Mexicans, and Popago Indians in Arizona. Oakland (1978) also found no

significant mean difference in Austin, Texas. These preliminary data suggest that the California SOMPA norms are holding up well in the Southwest and West. A second study by Reschley (1978) indicates through factor analysis that WISC-R subtest scores hold up across cultures on an intra-individual comparison basis.

All of the above studies offer indirect evidence that the SOMPA measures continue to be acceptable when used with groups of children outside of California.

The main focus of this paper is the use of the Estimated Learning Potential (ELP) as a non-discriminatory measure of intelligence. Two unpublished pieces of research will be presented (Mercer, personal communication, 1978) concerning this specific issue.

Oakland, in his study of the Austin, Texas study, used a sample of exceptional children, and constructed regression equations for the sample. These equations were sent to Mercer who rescored her California groups of Anglos, Blacks and Hispanics using Oakland's regression equation. The ELP's for the groups were changed by approximately three points. This change is statistically significant, but would make little practical difference.

Reschley (1978) in Pima County school district collected data on a stratified random sample of the school population (a representation of all children in the schools, both normal and exceptional). From this data a local regression equation was constructed and sent to Mercer. Mercer rescored her groups with Reschley's equation and found no significant difference with the original ELP scores.

All of the above studies strongly suggest that Mercer's published procedures for obtaining an ELP will hold up in Louisiana as they have in Arizona and Texas.

Required Procedures for Intellectual Assessment

The Board of Elementary and Secondary Education has approved the following provisions for non-discriminatory intellectual assessment under the requirements of Section 436 (C) of the Regulations implementing Act 754.

1. Every time an intellectual assessment is conducted the examiners shall assess the sociocultural characteristics of the child and the family using the Sociocultural Scales of the SOMPA.
2. When the child is 5 to 12 years of age and Anglo, Black, or Hispanic in ethnic background, and Estimated Learning Potential (ELP) shall be calculated and used as the indicator of intellectual abilities.
3. When the child is outside the age range or ethnic groups, the results of the standard intellectual measurement shall be interpreted by the psychologist taking the sociocultural results into account.